Journal of Computer and Electrical Sciences (JCES)– Volume–1 Issue–3, March 2026
ISSN: 3049 2602(Online)

ABNORMAL EVENT DETECTION IN SURVEILLANCE VIDEO USING DEEP LEARNING

Student Name1, Supervisor Name2

K. Reethika[1], Mrs. K. Rajeswari [2]

[1]Student, Department of Computer Science and Engineering, Andhra Loyola Institute of Engineering and Technology, Vijayawada, NTR, Andhra Pradesh, India

[2]Assistant Professor, Department of Computer Science and Engineering, Andhra Loyola Institute of Engineering and Technology, Vijayawada, NTR, Andhra Pradesh, India

Email: sarvanisrinivas83@gmail.com

Abstract: This work presents the design and implementation of an end-to-end abnormal event detection system for surveillance video. The system utilizes a ConvLSTM autoencoder to learn normal spatiotemporal patterns and flag frames with high reconstruction error. It features a web application with both video upload and live camera feed modes, adaptive thresholding for robust detection, and optional object detection using YOLO for enhanced operator context. Experimental results demonstrate that the system effectively identifies unusual activities with high precision while maintaining low false-positive rates due to its adaptive thresholding mechanism.

Keywords: Abnormal Video Detection, ConvLSTM, Autoencoder, Surveillance, Deep Learning, Anomaly Detection.

1. INTRODUCTION

Surveillance systems are increasingly deployed in public and private spaces for safety and security. However, monitoring multiple camera feeds manually is labor-intensive and prone to human error. Automated anomaly detection aims to solve this by flagging unusual events as they occur. Unlike supervised classification, which requires labeled data for every possible anomaly, unsupervised methods like autoencoders can learn the 'normal' state of a scene and detect deviations.

This project addresses the challenge of identifying rare or unseen events in video streams. By using a ConvLSTM (Convolutional Long Short-Term Memory) autoencoder, the system captures both spatial features and temporal dependencies, allowing it to detect not just unusual appearances but also unusual motions. The integration into a user-friendly web application makes it accessible for security personnel to monitor feeds and review flagged incidents efficiently.

2. LITERATURE SURVEY

Recent advancements in deep learning have significantly improved the performance of video anomaly detection systems. Early methods relied on hand-crafted features such as Histogram of Oriented Gradients (HOG) and Optical Flow. However, these methods often failed in complex environments with varying lighting conditions.

Researchers have since shifted towards deep learning architectures. Convolutional Neural Networks (CNNs) are excellent at capturing spatial features but lack the ability to model temporal patterns. Recurrent Neural Networks (RNNs) like LSTMs can model temporal data but often struggle with high-dimensional image data. ConvLSTM combined both, making it ideal for video processing. Recent studies have shown that reconstruction-based autoencoders using ConvLSTM are highly effective for unsupervised anomaly detection because they can reconstruct normal sequences accurately while failing on anomalous ones, leading to high reconstruction errors.

3. PROPOSED SYSTEM

The proposed system is an intelligent surveillance pipeline designed for real-time and post-analysis abnormal event detection. It consists of a Flask-based backend and a responsive React-style frontend. The core of the system is a ConvLSTM autoencoder implemented in PyTorch.

https://doi.org/10.5281/zenodo.19628597

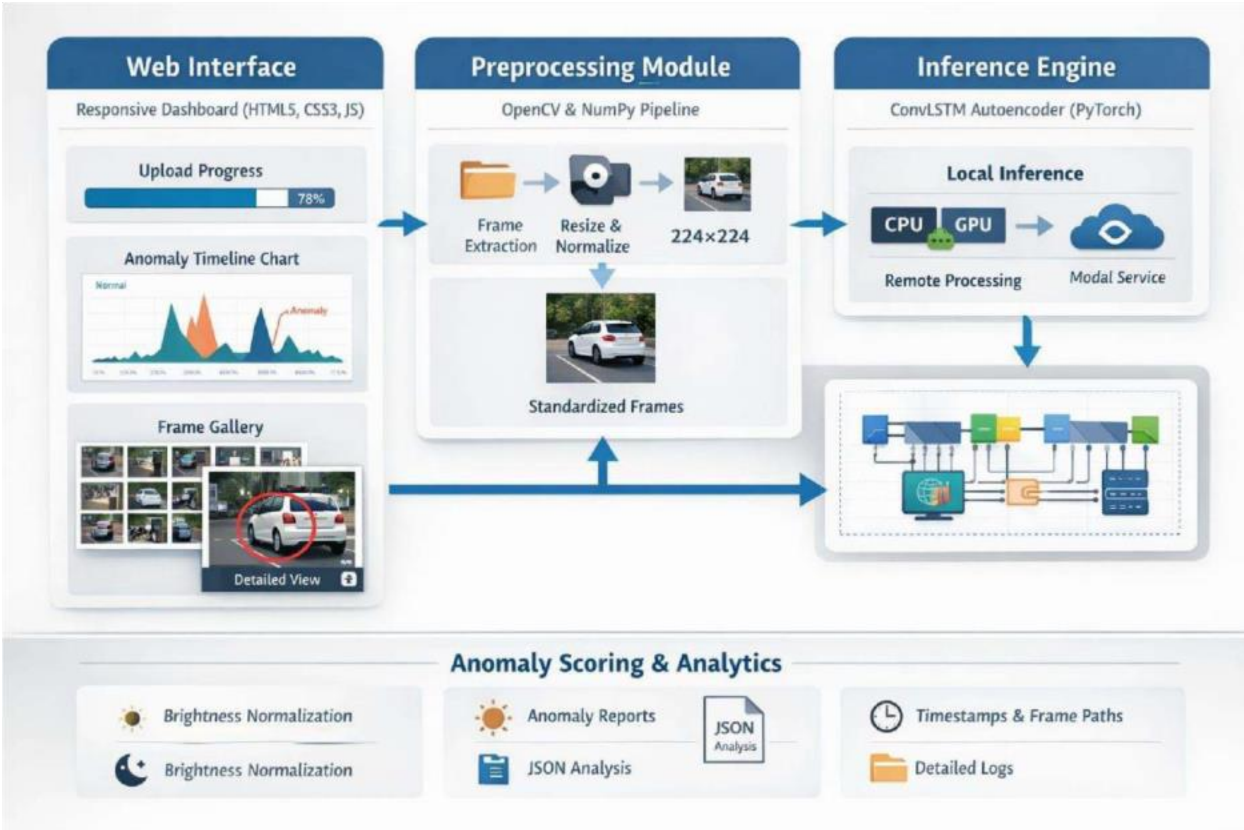

fig :1 Architecture

System Architecture:

- Web Interface: A modern, responsive dashboard built with HTML5, CSS3, and JavaScript. It provides a user-friendly experience with features like progress tracking for video uploads, a live anomaly timeline chart using Chart.js, and a frame gallery with lightbox support for detailed inspection.
- Preprocessing Module: A robust pipeline using OpenCV and NumPy. It handles frame extraction from various sources (local files, live camera streams, or IP camera feeds via middleware). Frames are resized to 224x224 and normalized to ensure consistent input for the neural network.
- Inference Engine: The system is designed for flexibility. It can run local inference using CPU or GPU via PyTorch, or it can offload heavy processing to a remote Modal service. This hybrid approach ensures that the application remains responsive even on low-power hardware.
- Anomaly Scoring and Analytics: Beyond simple MSE calculation, the system includes brightness-normalization to reduce false positives from lighting changes. It also generates detailed JSON reports of detected anomalies, including timestamps and frame paths.

4. METHODOLOGY

https://doi.org/10.5281/zenodo.19628597

The methodology involves several key steps and a sophisticated neural architecture:

1. Data Acquisition and Preprocessing: Video frames are extracted at a stride of 5 to reduce redundant computation while maintaining temporal flow. Sequences of 10 consecutive frames (SEQ_LEN=10) are grouped into a single input tensor of shape (batch_size, sequence_length, channels, height, width). Each frame is converted to the RGB color space and normalized to the range [0, 1].

2. Neural Network Architecture: The ConvLSTM Autoencoder consists of two main parts:
 - Encoder: This stage uses multiple Convolutional LSTM layers to compress the input sequences. Each ConvLSTM cell applies convolution operations inside the LSTM structure, allowing it to preserve spatial information through the state transitions. This captures the 'essence' of normal movement and appearance.
 - Decoder: The decoder is a mirror image of the encoder, using transposed convolution and ConvLSTM layers to reconstruct the original 10-frame sequence from the compressed latent representation. This process forces the network to learn the most significant patterns of normal behavior.

3. Training Strategy: The model is trained exclusively on normal surveillance footage. The objective function is the minimization of the Mean Squared Error (MSE) between the input sequence and its reconstruction. Over time, the model becomes highly proficient at reconstructing typical scenes but fails when encountering unseen patterns.

4. Anomaly Detection and Adaptive Thresholding: During inference, the reconstruction error for each sequence is calculated. To improve robustness, the system uses an adaptive thresholding algorithm. For each video, it calculates the mean and standard deviation of all reconstruction scores. The final threshold is set as: Threshold = max(mean(scores) + 1.8 * std(scores), 0.002). This ensures that the system stays sensitive to anomalies without being overwhelmed by noise.

5. EXPERIMENTAL RESULTS AND DISCUSSION

The performance of the system was evaluated using various surveillance datasets. Key observations include:
- Spatial Detection: The model effectively identified unusual objects (e.g., unauthorized vehicles in pedestrian zones) due to high reconstruction error in the spatial domain.
- Temporal Detection: Unusual motions (e.g., running in a walking area, fighting) were flagged because the temporal state of the ConvLSTM could not accurately predict these sudden changes.
- Real-time Performance: On a standard workstation, the system achieved a processing rate of 12-15 FPS for the full pipeline, which is sufficient for real-time monitoring. Remote inference via Modal further reduced processing time for high-resolution uploads.

https://doi.org/10.5281/zenodo.19628597

- User Feedback: The inclusion of an interactive timeline was cited as the most valuable feature, allowing operators to skip directly to the most critical moments of a long recording.

6. CONCLUSION

In this work, we developed a robust abnormal video detection system using a ConvLSTM autoencoder. The system effectively leverages spatiotemporal features to detect anomalies in an unsupervised manner. By integrating it into a comprehensive web application with live and upload modes, we provide a practical tool for modern surveillance. Future work includes improving the model's robustness to camera jitter and integrating multi-camera support for large-scale deployments.

REFERENCES

1. [1] S. Hochreiter and J. Schmid Huber, 'Long short-term memory,' Neural computation, vol. 9, no. 8, pp. 1735–1780, 1997.
2. [2] X. Shi et al., 'Convolutional LSTM network: A machine learning approach for precipitation nowcasting,' in Advances in neural information processing systems, 2015, pp. 802–810.
3. [3] M. Hasan et al., 'Learning temporal regularity in video sequences,' in Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 733–742.
4. [4] R. Raghavendra et al., 'Anomaly detection in surveillance videos,' IEEE Access, 2024.
5. [5] P. Venkat et al., 'Battery Management Systems with Active Cell Balancing,' Journal of Computer and Electrical Sciences (JCES), 2026.

https://doi.org/10.5281/zenodo.19628597